7,5/7,5

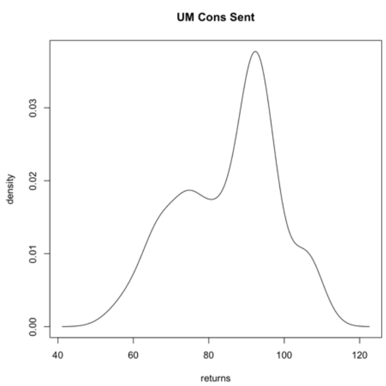
Steve Depp
413 – 55

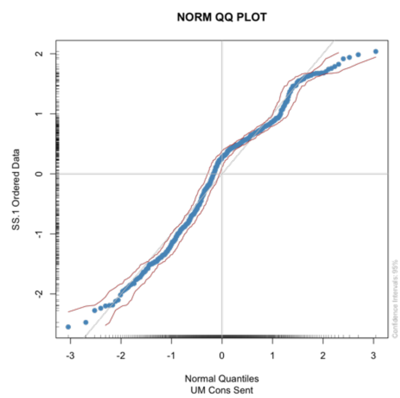
Assignment 3
15 July 2019

1.1 The following presents sample size, mean, standard deviation, median, 20% trimmed mean, min, max, range, skewness, excess kurtosis for the monthly series of Consumer Sentiment of the University of Michigan from January 1, 1978 to August 1, 2013. First, I confirmed there are no gaps in the 428 monthly observations and all dates are 1st of the month. Dates are then decimalized dates now = year + months / 12.

	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
VALUE	428	85.21565	13.14239	88.85	85.52674	51.7	112	60.3	-0.310136	-0.6886821

The **above** univariates describe a distribution departing from normal via moderate negative skew with moderate shoulders, negative kurtosis. Trimmed mean > median > mean and a minimum extending more than half the range below center are consistent with negative skew, though trimmed so close to mean would lead one to think the distribution is skewed negatively inside the middle 80% of the distribution.

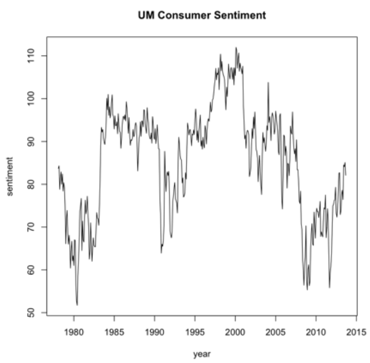


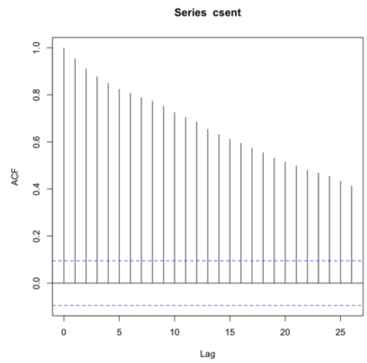


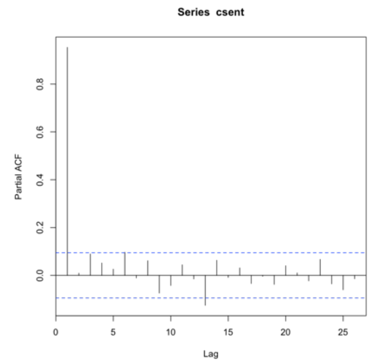
The density plot **above** has 2 modes, one obviously more prominent than the other, which explain the negative skew arising from the distribution midsection, negative kurtosis arising from the left shoulder / 2nd mode. What is notable here (and nowhere else) is that the mean is the least likely of observations inside the interquartile range (actually further than 75th percentile on the RHS. 25th percentile is 74.475 and 75th is 94.225.

The reduced density at the mean can be seen where the **above** QQ plot mean crosses to the left of the 0th normal quantile. Excluding 7 – 8 fringe points, the distribution demonstrates negative kurtosis on its RHS not LHS. Notice the data extends 2++ SDs to the left (bottom of the y-axis) roughly equivalent to the normal distribution, but less than 2 SDs to the right (top of the y-axis) where normal would be 3 SD. Otherwise the distribution appears here relatively normal, rendering the density plot very useful.

A test of null hypotheses that the distribution is symmetric with respect to the mean yields a test statistic = -2.62 and p-value = 0.009 allowing us to reject the null hypothesis with greater than 99.9% level of confidence. A similar test for kurtosis yields a test statistic -2.91 and p-value 0.004 allowing us to reject the null hypothesis that excess kurtosis = 0 with greater than 99.9% level of confidence.







The time series plot **above** shows sharply changing levels, and no trend obvious across the whole data. No consistent change or progression in variance is evident across the range though variance does vary in the data and in ranges of the data. So, log transform appears unneeded and un-useful. No discernable seasonality from this distance. Some discernable pattern to cyclicality but nothing to depend upon: 7 years sideways with 15% ranges then shifts lower. Shorter sideways periods (~5 years) with larger ranges and shifts upward. These patterns are aperiodic and unreliable.

The ACF plot **in center above** with steadily declining autocorrelation levels is consistent with the single near unity partial correlation at lag = 1 in the PACF plot on the **RHS above** and suggests that a 1st difference would likely remove autocorrelations from both plots. The significantly negative partial autocorrelation at lag = 13 should be investigated for whether its appearance is consistent with the randomness of a 95% confidence interval or truly represents some form of 13-month seasonality.

```
[1] "constant w/ no trend"


Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -2.2675
  P VALUE:
    0.2129
```

```
[1] "constant w/ trend"


Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -2.2638
  P VALUE:
    0.4661
```

```
[1] "no constant or trend"


Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -0.305
  P VALUE:
    0.5191
```

There are several tests for unit roots to apply to this data. **Firstly**, the Augmented Dickey Fuller Test's null hypothesis is that there are unit roots and the data is not stationary. The alternate hypothesese are that the series is stationary or trend stationary (1) with a constant but no trend (type = "c") that appears to fit our raw data, (2) with neither an intercept nor trend (type = "nc") or (3) with intercept and trend (type = "ct"). Testing the null hypothesis with a lag = 5 allows us to reject the null hypothesis under all the above alternatives and tests for unit roots for models within a lag of 5 or shorter, thus incorporating the 1st difference lag we investigate in problem 2. In all 3 cases **shown above**, the pvalue is greater than 0.05 and we are unable to reject the null hypothesis that our data is non-stationary and has unit roots. **Secondly**, KPSS Unit Root Test is another test which automatically employs lag = 5 for our data set and tests the null hypothesis that our data are stationary. Here the results yield a test statistic 0.8935 > critical value = 0.739 which allows us to reject the null hypothesis that the data is stationary at the 99% confidence level. **Finally**, KPSS tests can be repeatedly executed via the 'ndiffs' function to determine the appropriate number of first differences, which in this case is 1.

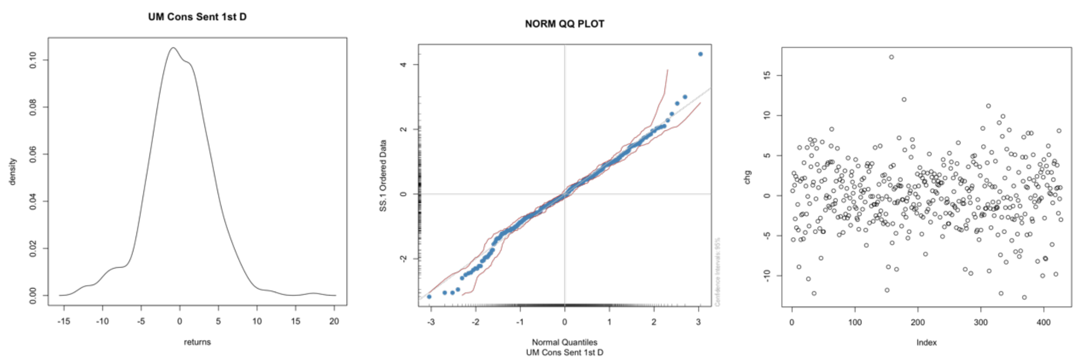
```
        One Sample t-test

data:  chg
t = -0.019349, df = 426, p-value = 0.9846
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.3843808  0.3768867
sample estimates:
   mean of x
-0.003747073
```

⟹ no linear trend

2.1 As shown above, the sample mean of 427 first differences across our 428 observations equals -0.0037 which is captured within a 95% confidence interval containing zero with test statistic = -0.019349 and p-value = 0.9846, that do not allow us to reject the null hypothesis that the mean or expected rate of change in UMich consumer sentiment is zero. To employ this test, we would need some sense of the distribution of 1st differences, which is presented here:

	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
X1	427	-0.003747073	4.001635	-0.2	0.04110787	-12.7	17.3	30	-0.05806149	1.111104

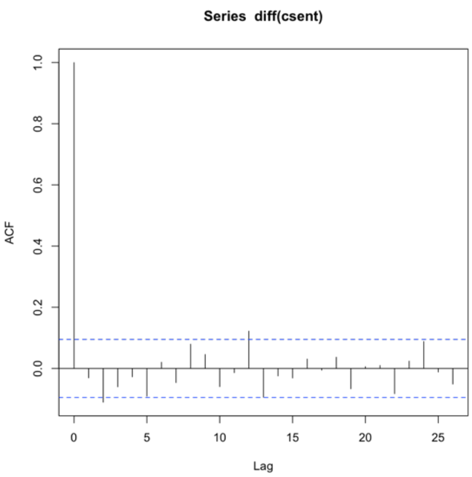
First differences' configuration of median < mean < trimmed mean shown **above** is consistent with its moderate negative skew, though the distribution of the range appears skewed to the topside: max is 17.1 above the median and min is only 12.9 below.



It would appear from the **above** plots that the excess kurtosis which is distributed on the RHS more so than LHS is responsible, though the pointy mode which is also leftward leaning may also be to blame. The QQ plot also **above** provides corroborating evidence as the densest part of the distribution extends 2.5 SDs to the left and only 2 SDs to the right, while 4 points extend another full SD to the RHS (filling out 3SDs) and a 5th extends the RHS of the distribution another 1++ SDs to 4 SDs from middle where normal would find this point only 3 SDs from middle. On the LHS, 4 points fill out 3 SD of our distribution more or less in line with normal. The time series plot **above** appears to be randomly distributed with mean zero.

The null hypotheses of an unskewed distribution cannot be rejected based upon a test statistic = -0.49. On the other hand a null hypothesis of no excess kurtosis can be rejected with a test statistic = 4.68 with greater than 99.9% confidence.

Overall thus, versus the raw data, 1st differences has switched from bimodal to single modal, thin to fat tails, negative to zero skewed. Net, the 1st differences distribution is more suitable for modeling.




```
Box-Ljung test

data:  chg
X-squared = 24.111, df = 12, p-value = 0.01965
```

```
[1] "no constant or trend"
Warning message in adfTest(chg, lags =
"p-value smaller than printed p-value"

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 12
  STATISTIC:
    Dickey-Fuller: -6.0438
  P VALUE:
    0.01

Description:
 Mon Jul 15 17:52:22 2019 by user:
```

2.2 The ACF of the 1st differences implicitly carries out multiple hypothesis tests, each once with alpha/2 probability of false positive. When complete it is thusly likely to have 1 false positive. Ljung-Box tests the first h autocorrelations as a group for significant departures from zero, such that the null hypothesis with h=12 is correlation_1 = correlation_2 = … = correlation_12 = zero. **Above** the results of Box-Ljung test have a p-value of 0.01965 indicating we can reject the hypothesis of zero correlations from lag = 1 to 12. (My guess is that running a Dickey-Fuller Test to lag 12 would provide similar confidence that there are no unit roots available out to lag 12. **Above**, I run an adfTest settting type="nc" to obtain a pvalue less than 0.01, which is consistent with rejecting the null hypothesis that the 1st differenced data is stationary and has no unit roots.)

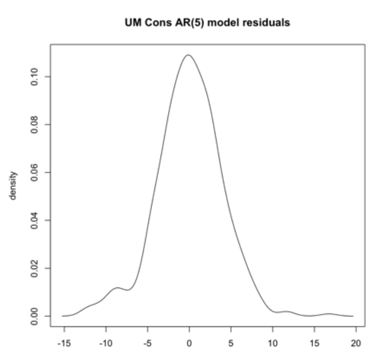
```
Call:
ar(x = chg, method = "mle")

Coefficients:
      1        2        3        4        5
-0.0499  -0.1266  -0.0830  -0.0498  -0.1107

Order selected 5  sigma^2 estimated as  15.46
```

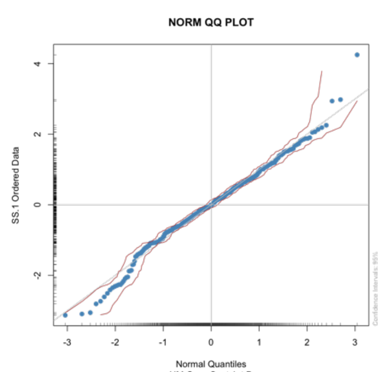
```
Call:
arima(x = chg, order = c(5, 0, 0), include.mean = F)

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -0.0499  -0.1266  -0.0830  -0.0498  -0.1107
s.e.   0.0481   0.0481   0.0483   0.0483   0.0483

sigma^2 estimated as 15.46:  log likelihood = -1190.59,  aic = 2393.17
```

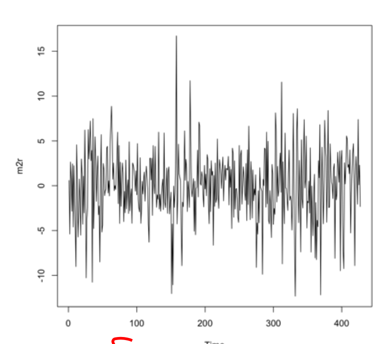
3.1 Both calls produce the same model for the 1st differences data. The arima() on RHS is a specified model and ar() on the LHS is optimized based upon AICc. The model is:

$$\mathbf{y_t = -0.0499*y_{t-1} - 0.1266*y_{t-2} - 0.0830*y_{t-3} - 0.0498*y_{t-4} - 0.1107*y_{t-5} + \varepsilon_t}$$

3.2 Model validation plots **below** closely resemble those of 1st differences examined above.

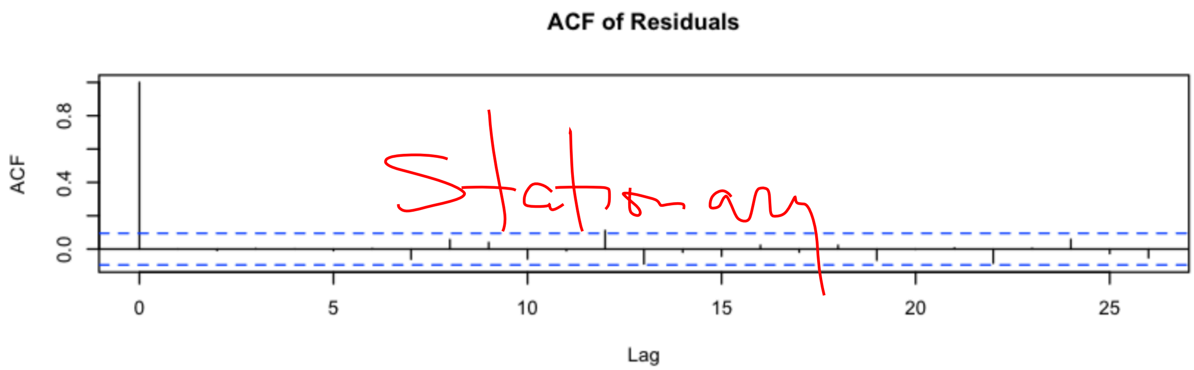


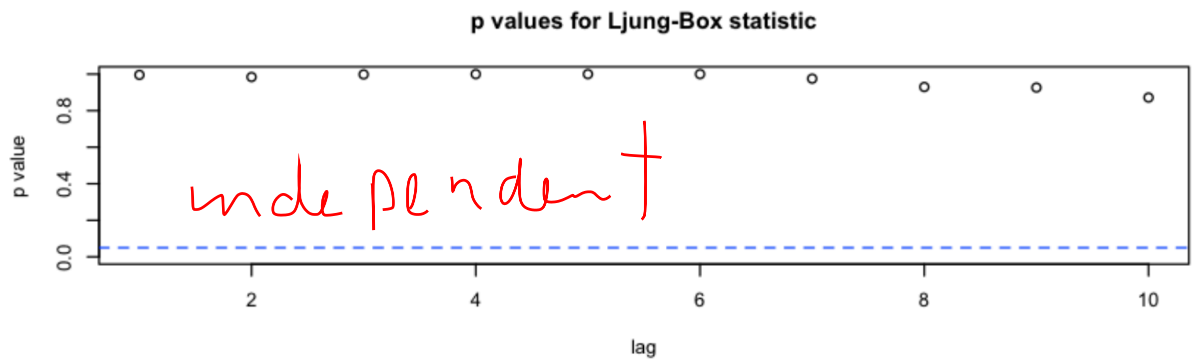




	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
X1	427	-0.004996033	3.936784	-0.03631819	0.06054516	-12.3328	16.72234	29.05514	-0.1155215	1.104878

symmetric about zero






```
        Ljung-Box test

data:  Residuals from ARIMA(5,0,0) with zero mean
Q* = 5.2815, df = 5, p-value = 0.3825

Model df: 5.   Total lags used: 10
```

ACF and Ljung-Box results suggest residual autocorrelation, un-modeled relationships between lagged observations of the residuals, are not significant. The p-values for every lag in the 2nd plot **above** indicate we must fail to reject the null hypothesis that the residuals are stationary, that jointly the correlations are zero at the 95% confidence level. The specific Ljung Box test below it agrees for the specific lag = 5.

`[1] "student t"`

ar1	-1.0374207878589
ar2	-2.63102526856085
ar3	-1.71749358634343
ar4	-1.03237780926274
ar5	-2.2933074899817

`[1] "p value of coefficient != zero"`

ar1	0.300135269781214
ar2	0.00882520921038865
ar3	0.0866244013858091
ar4	0.302487851094867
ar5	0.0223223526062906

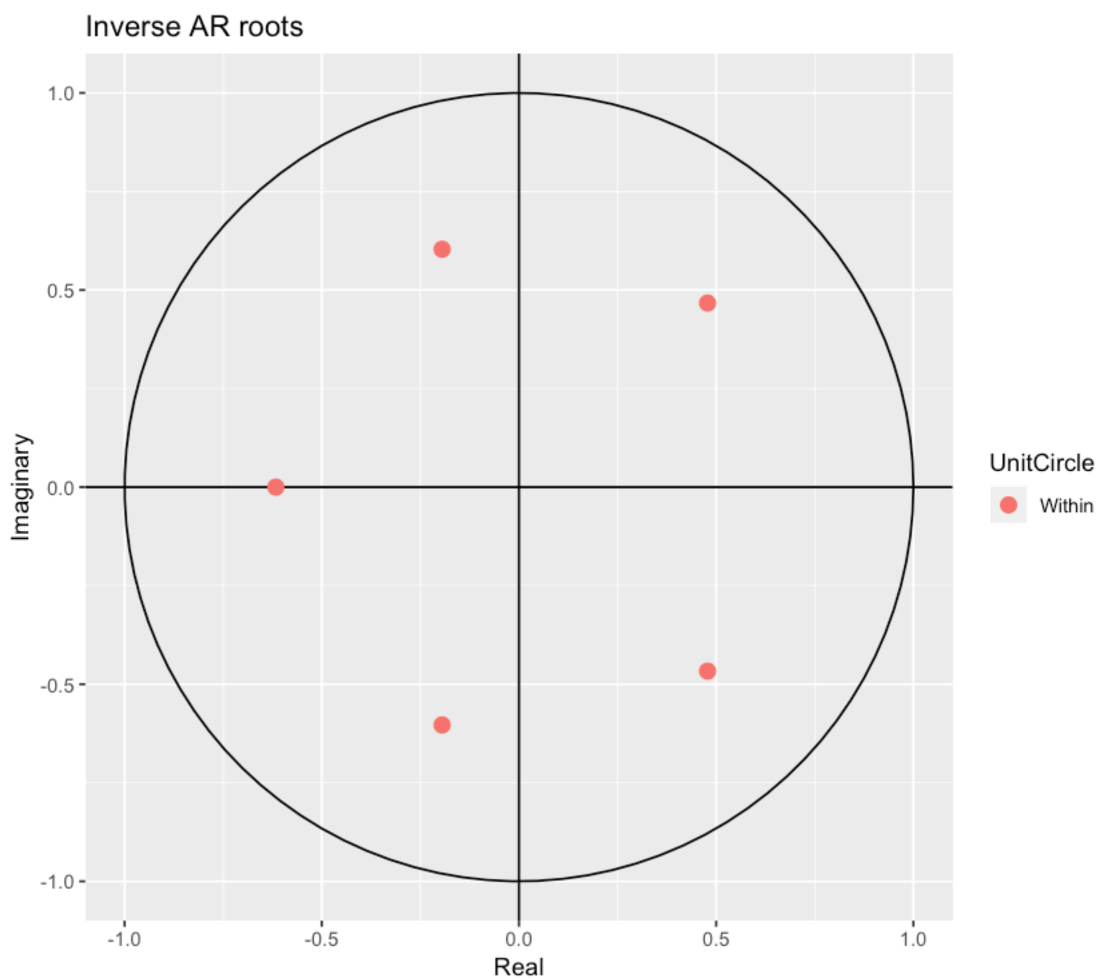
To evaluate coefficients' significance, we can divide their absolute value magnitude by standard error and compare with a student's t distribution. Above 1.2 would be significant at the 95% confidence level. (Alternatively, I read that since arima optimizes via maximum likelihood, the coefficients are asymptotically normal enabling us to use z-statistics to calculate p-values.). The student-t pvalues are listed **above** suggesting that the lag 1 and lag 4 parameters are not significant.

3.3 The business cycles in consumer sentiment are aperiodic, unpredictable, or not predictable enough to betray stationary nature of the 1st differences. Yet this irregular cycle to the data, which I mentioned in the EDA segment, can be averaged and modeled via sin and cosine waves, the imaginary part of complex unit roots to the model. For this

AR(5) model the 5 roots are composed of one real and 2 pairs of polar complex roots shown in the plot **below**: The inverse of the real root is -1.622 shown below inside the unit circle to retain invertibility = 1/–1.622 = –0.617 along the x-axis. The other 2 roots inverses are complex -0.485 +/– 1.501i and 1.071 +/– 1.045i whose imaginary parts show up **below** as +/– 0.755 and +/– 0.957 on the imaginary or y-axis. For practical use, these imaginary roots in the form a+bi enter the following formula:

$$k = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2+b^2})}.$$

to compute average cycle length = 8.124 and 3.336 in the 1st differences model which is incremented by 1 for original terms of UM consumer sentiment data ~ 9.1 months and 4.4 months average cycle length, each cycle working in opposite directions in sin and cosine fashion.

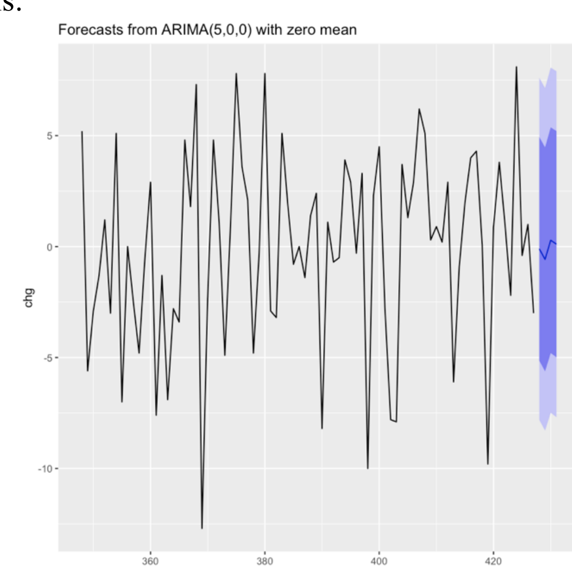


```
$pred
Time Series:
Start = 428
End = 431
Frequency = 1
[1] -0.1038796 -0.5743992  0.2851835  0.1059215

$se
Time Series:
Start = 428
End = 431
Frequency = 1
[1] 3.932175 3.937062 3.967210 3.976876
```

3.4 Remember that with 1st differences, we have 427 observations for 428 original measurements of UM consumer sentiment. The 1 to 4 steps ahead prediction within the scale of 1st differences provides 1st differences for future observations 428 = -0.1039, 429 = -0.5744, 430 = 0.2852 and 431 = 0.1059 with associated standard errors as shown in the output **above**. The same predictions of 1st differences commencing with the difference between Sep 2013 and Aug 2013 and continuing through the 1st difference between Dec 2013 and Nov 2013 is shown **below** together with the confidence band = the point estimate + or – the standard error * 1.96. You can see the band is quite wide particularly for differences but I suppose we need to keep in mind the variance is 15.46 for equally scaled standard deviation of 3.93. So, the width of the band is about 4 times the standard deviation of the model.

dates	S-A13	O-S13	N-O13	D-N13
lcl	-7.8109	-8.291	-7.4905	-7.6888
fcast	-0.1039	-0.5744	0.2852	0.1059
ucl	7.6032	7.1422	8.0609	7.9006

The **above** calculates these figures by hand. The model provides an automatic plot and table **below** with 80 and 95% confidence intervals:




```
    Point Forecast      Lo 80    Hi 80      Lo 95    Hi 95
428     -0.1038796 -5.143164 4.935405 -7.810800 7.603041
429     -0.5743992 -5.619948 4.471149 -8.290899 7.142101
430      0.2851835 -4.799001 5.369368 -7.490406 8.060773
431      0.1059215 -4.990650 5.202493 -7.688612 7.900455
```

4.1 Reprinting **below** the student-t values and p-values for the coefficients of the model, we will remove those coefficients with high p-values, ar1 and ar4, which represent the lag = 1 and lag = 4 parameters. The resulting model is shown to the right and retyped **below:**

```
[1] "student t"

ar1   -1.0374207878589
ar2   -2.63102526856085
ar3   -1.71749358634343
ar4   -1.03237780926274
ar5   -2.2933074899817

[1] "p value of coefficient != zero"

ar1   0.300135269781214
ar2   0.00882520921038865
ar3   0.0866244013858091
ar4   0.302487851094867
ar5   0.0223223526062906
```

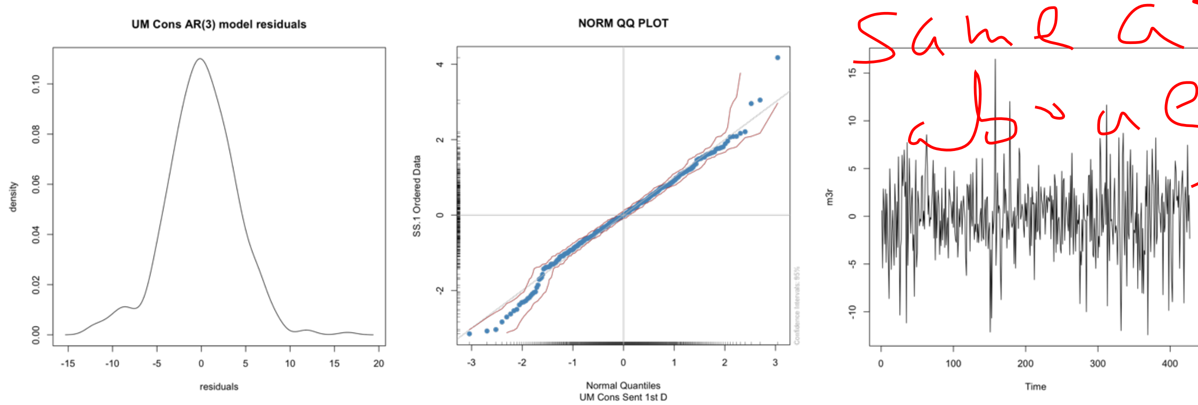
```
Call:
arima(x = chg, order = c(5, 0, 0), include.mean = F, fixed = c1)

Coefficients:
      ar1      ar2      ar3  ar4      ar5
        0  -0.1191  -0.0754    0  -0.1067
s.e.    0   0.0479   0.0480    0   0.0483

sigma^2 estimated as 15.54:  log likelihood = -1191.59,  aic = 2391.17
```

$$y_t = -0.1191*y_{t-2} - 0.0754*y_{t-3} - 0.1067*y_{t-5} + \varepsilon_t$$

The model coefficients and standard errors and overall standard deviation does not change much at all. AIC is 2 lower from 2393.17 to 2391.17, due to fewer 2 fewer parameters.

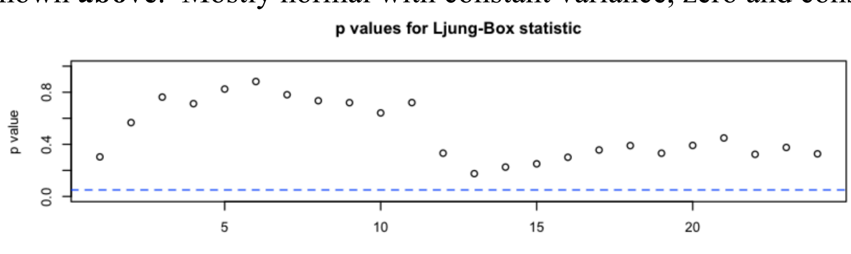
	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
X1	427	-0.004376082	3.946085	-0.009550305	0.06094165	-12.42267	16.46735	28.89002	-0.1143944	1.064369

4.2 I would think the AR(3) model has similar residual characteristics to previous AR(5) given the similar parameters and performance. None of the univariate measures above differ much from AR(5):



Same as above

None of the AR(3) plots designed for normality checks differ much from AR(5) either as shown **above**. Mostly normal with constant variance, zero and constant mean.



What is notable and significant is that the plot of Ljung-Box autocorrelation tests, shown **above**, has moved from rejecting the null hypothesis of zero correlations to nearing on rejecting the null hypothesis at a lag = 13. All other autocorrelation p-values are significantly lower as well. None of this should surprise as this decline in p-values indicates that the AR(3) residuals contain some autocorrelation that is not modeled by AR(3) because we have left 2 parameters out of the AR(3) model that were in the AR(5) model, whose residuals failed to reject the Ljung-Box null hypothesis (pvalues all near 99%).

Net, it appears that the AR(3) model is adequate, but not capturing as much autocorrelation as the AR(5) version. In terms of magnitude, both models have parameters which are sizably away from zero considering they are 1st difference predictors. So there would be no issue keeping either on that basis.

4.3 The AR(3) model provides a lower AIC than AR(5) as mentioned above. I would prefer AR(3) for fit and simplicity.

4.4 The business cycles in AR(3) are modeled slightly longer from 8.124 to 8.515 months from 3.466 to 3.336 months in 1st differences. I notice that the real parameter has switched from 5th place to 2nd place in the listing of unit roots. I am not sure if this is switching from the 5th lag to the 2nd lag or not.

AR(3) unit roots:

```
1.15775100438727+1.05264341844673i   -1.56526912536364-0i   -0.37511644170545-1.51801062546045i
1.15775100438727-1.05264341844673i   -0.37511644170545+1.51801062546045i
```

AR(5) unit roots:

```
1.07069764140681+1.04541104793779i   -0.4846969466695+1.50086527242908i   -0.4846969466695-1.50086527242908i
1.07069764140681-1.04541104793779i   -1.62216202268765+0i
```

4.5 Models AR(3) and AR(5) return the exact same back-test results regardless of start date (300, 380, 420, etc). They are similar models, but I would not expect this to be the case. Including the screen shot just in case I have done something incorrectly here. I altered some code to experiment with the ‘fixed’ parameter in the function, but was unable to get this to work. I have gone back to think about this a lot more but cannot resolve except to suggest that h=1 might indicate that we are predicting based upon the same information regardless of the lags employed.

```
m2

Call:
arima(x = chg, order = c(5, 0, 0), include.mean = F)

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -0.0499  -0.1266  -0.0830  -0.0498  -0.1107
s.e.   0.0481   0.0481   0.0483   0.0483   0.0483

sigma^2 estimated as 15.46:  log likelihood = -1190.59,  aic = 2393.17
```

```
backtest(m2,chg,380,1,inc.mean=T)
```

```
[1] "RMSE of out-of-sample forecasts"
[1] 4.002229
[1] "Mean absolute error of out-of-sample forecasts"
[1] 3.120026
```

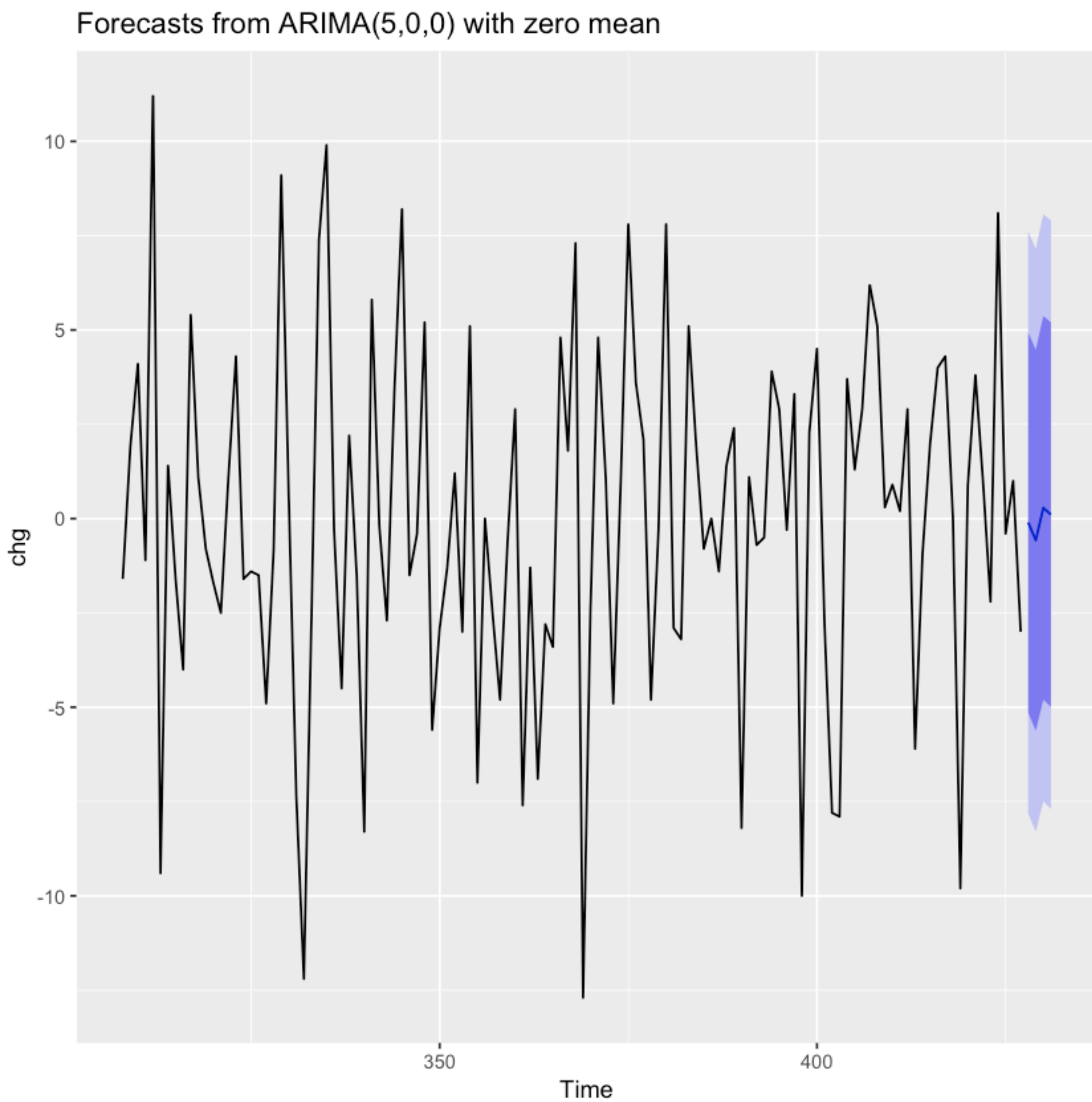
```
m3

Call:
arima(x = chg, order = c(5, 0, 0), include.mean = F, fixed = c1)

Coefficients:
      ar1      ar2      ar3  ar4      ar5
        0  -0.1191  -0.0754    0  -0.1067
s.e.    0   0.0479   0.0480    0   0.0483

sigma^2 estimated as 15.54:  log likelihood = -1191.59,  aic = 2391.17
```

```
backtest(m3,chg,380,1,inc.mean=T)
```

```
[1] "RMSE of out-of-sample forecasts"
[1] 4.002229
[1] "Mean absolute error of out-of-sample forecasts"
[1] 3.120026
```

5. The best model is the one which combines adequacy with fit and simplicity. Parameter count alone improves fit = AIC and simplicity for the AR(3) model over AR(5), but explaining a model that prescribes "lag every month to 5" might be easier than "lag just the 2nd, 3rd, and 5th months". On those bases, I would choose AR(5). Yet, something tells me that the Ljung Box high p-values might indicate overfit and that an AR(3) model might generalize better.

Executive report: We propose a model to predict UMich consumer sentiment which forecasts month over month changes in the index based upon reports from 2 months ago, 3 months ago and 5 months ago. On the basis of this model, we expect the index to continue to have an average monthly change of zero but with 95% confidence, we can narrow that forecast range from the +/- 10 shown in the plot below to +/- 7 index points shown in the forecasted light purple region. Furthermore, we note that the index responds to two opposing cyclical forces that last on average 3 and 8 months long.



n see from this plot

need this in natural units